

# Student project: Gene networks via PascalX

Daniel Krefl

Computational Biology Group

[daniel.krefl@unil.ch](mailto:daniel.krefl@unil.ch)

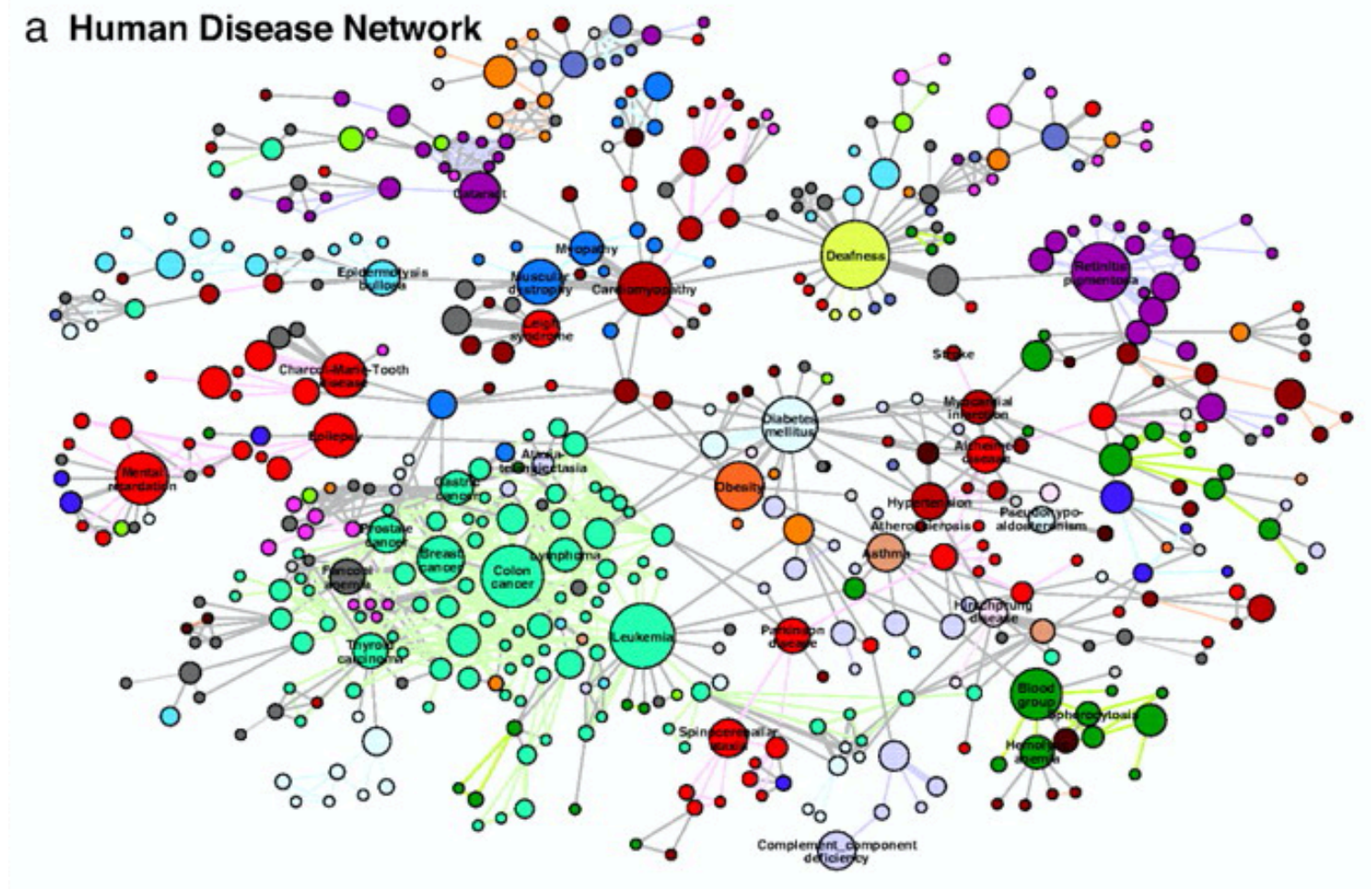
# Motivation



## Human disease network

( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:



( from Goh et al PNAS '07: The human disease network )



# Motivation

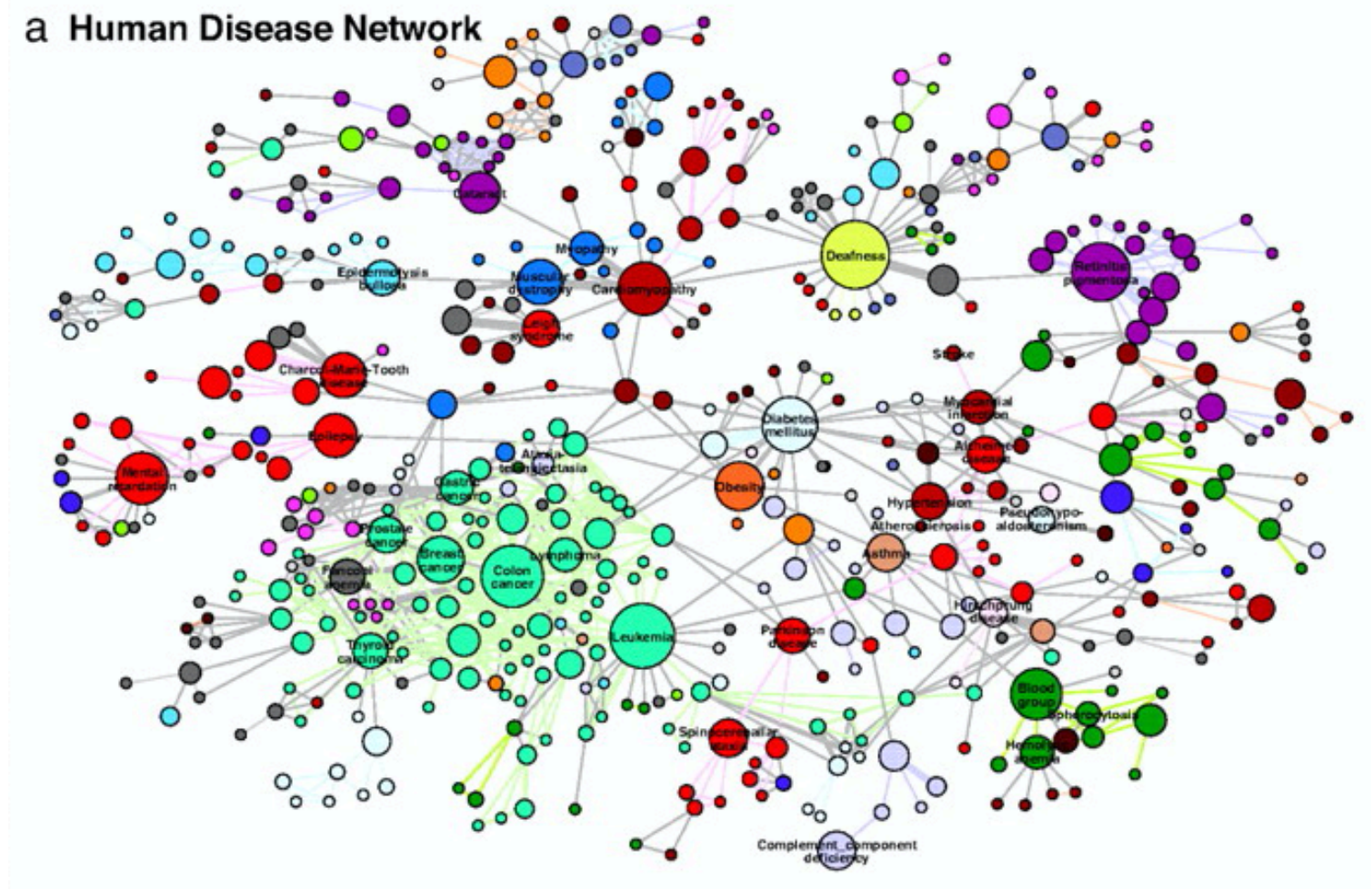
## ★ Human disease network

( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ Usually constructed via:

- Expert knowledge
- Textual data mining



( from Goh et al PNAS '07: The human disease network )

# Motivation

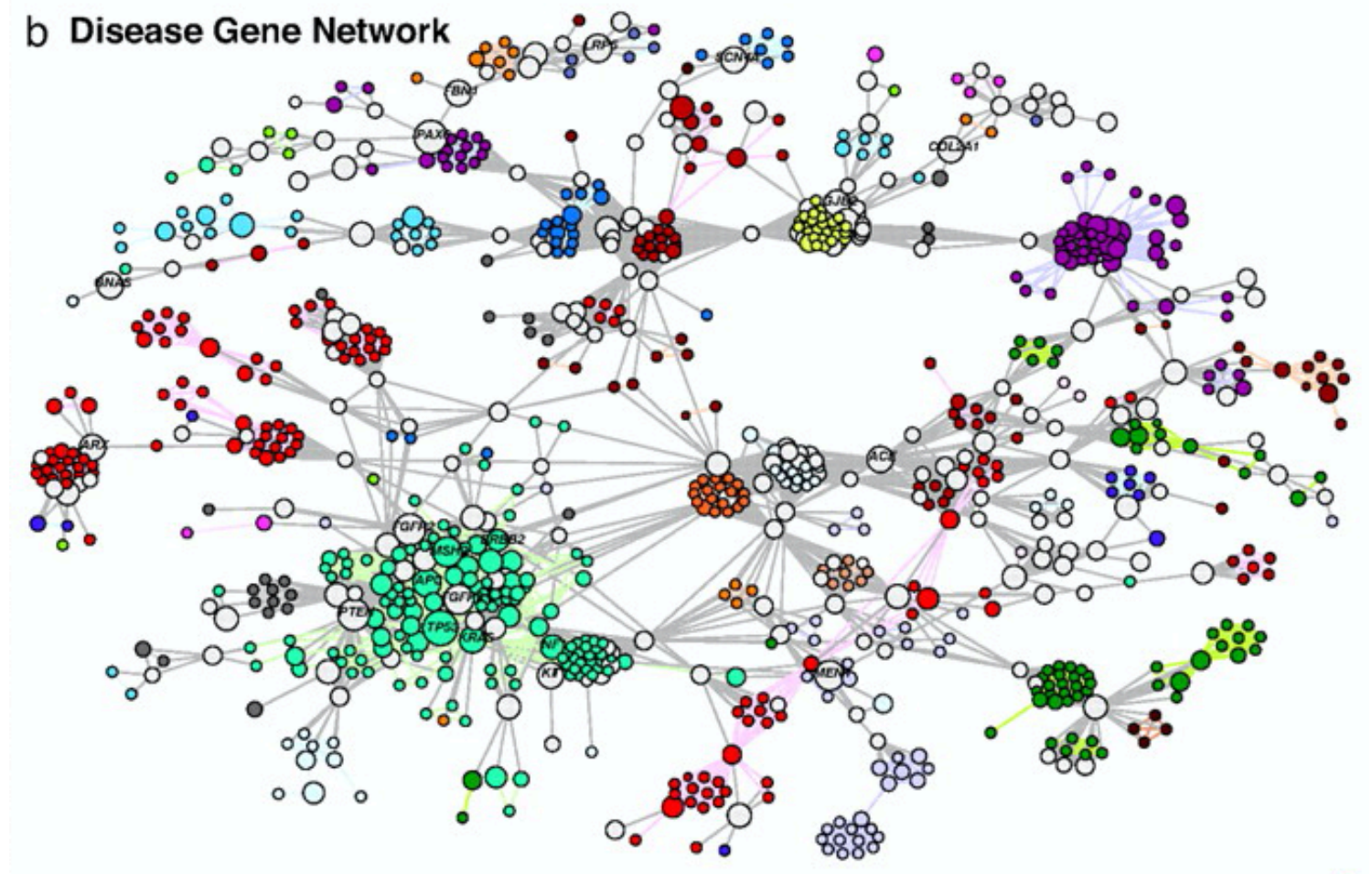
## ★ Human disease network

( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ Usually constructed via:

- Expert knowledge
- Textual data mining



( from Goh et al PNAS '07: The human disease network )



# Motivation

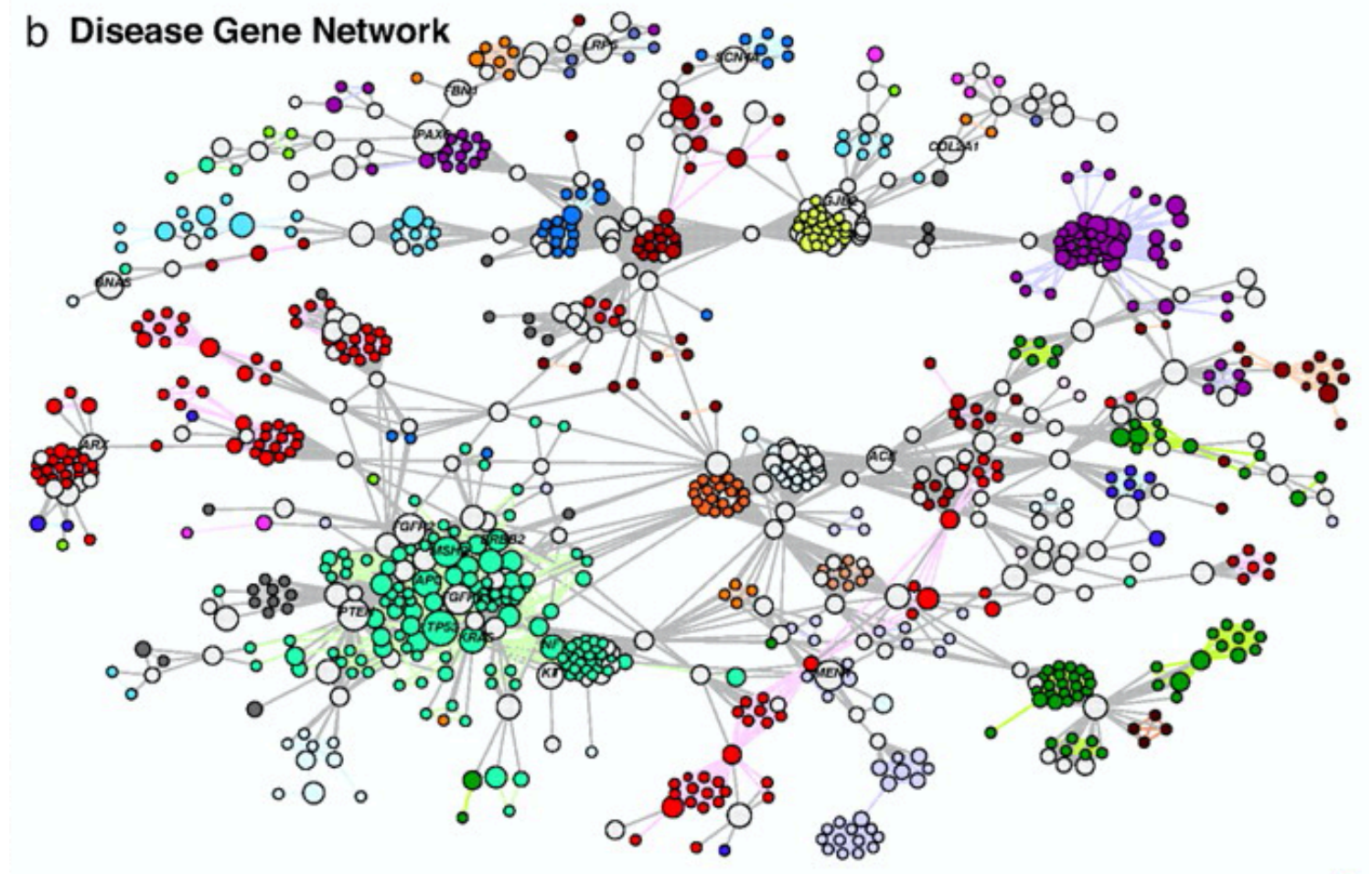


## Human disease network

( Goh et al '07: The human disease network )

Such networks yield novel insights:

- Shared genetic origins of diseases
- Multi-pathway association of diseases
- Disfunction of shared functional modules



( from Goh et al PNAS '07: The human disease network )

# Motivation



## Human disease network

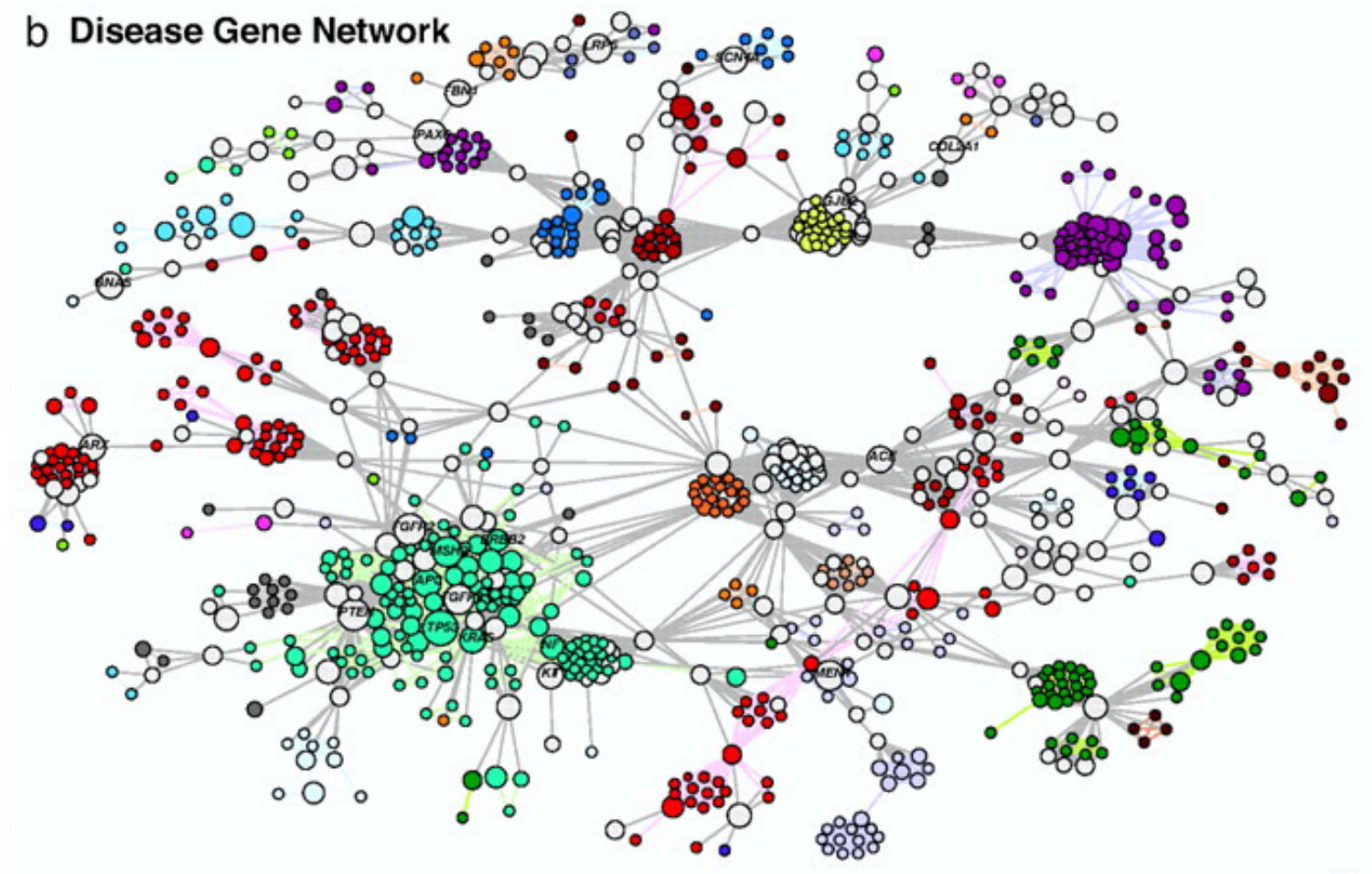
( Goh et al '07: The human disease network )

Good first read:

Bauer-Mehren et al.,  
Gene-disease network analysis reveals functional  
modules in mendelian, complex and environmental  
diseases

### Applications:

- Predict adverse drug reactions
- Predict new disease related genes
- Drug repurposing
- Identify shared mechanisms



( from Goh et al PNAS '07: The human disease network )



# Project

## ★ Human disease network

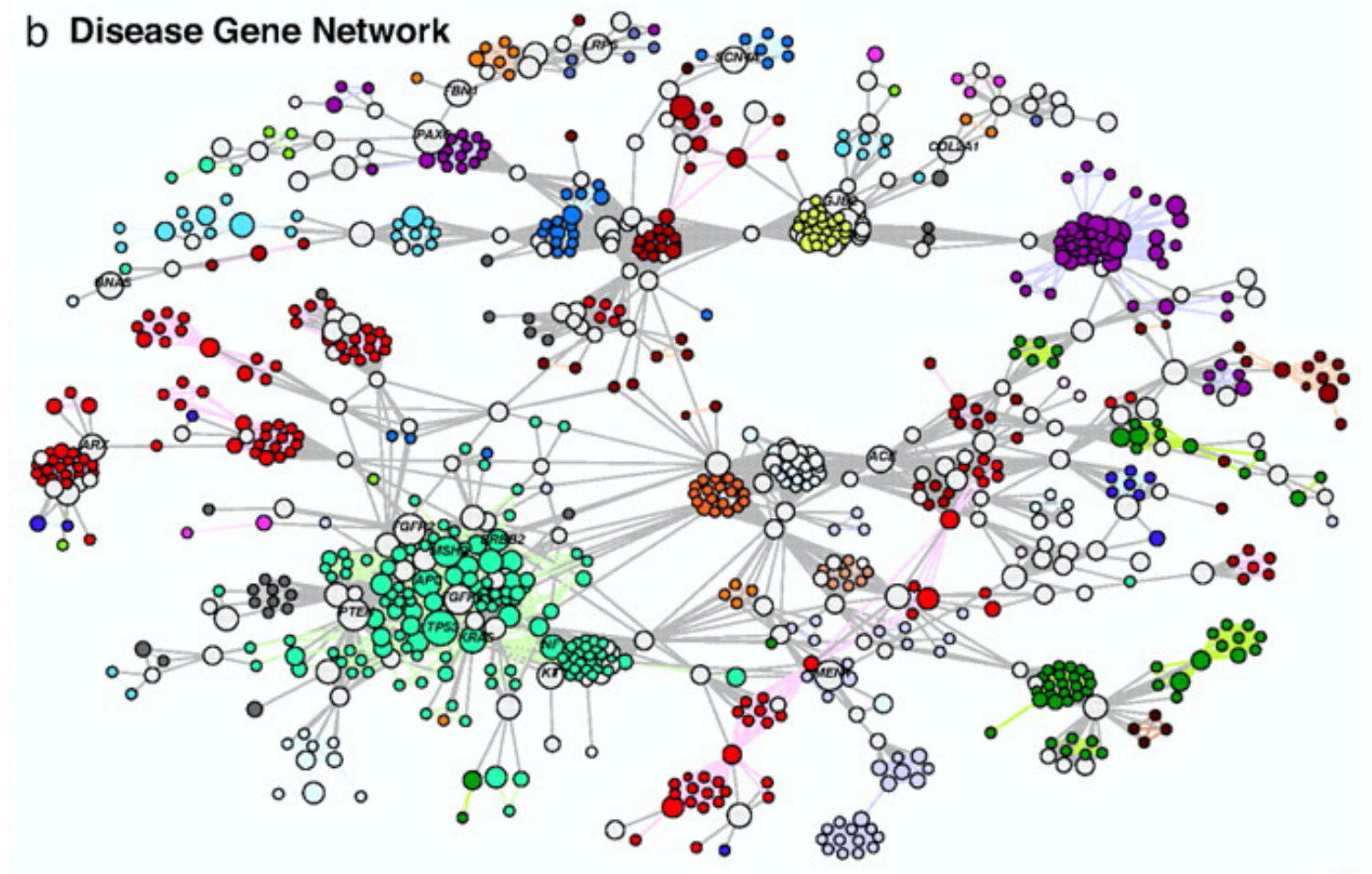
( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ Problems:

Construction via text mining:

- High effort
- Very noisy
- Not constant



( from Goh et al PNAS '07: The human disease network )

# Project

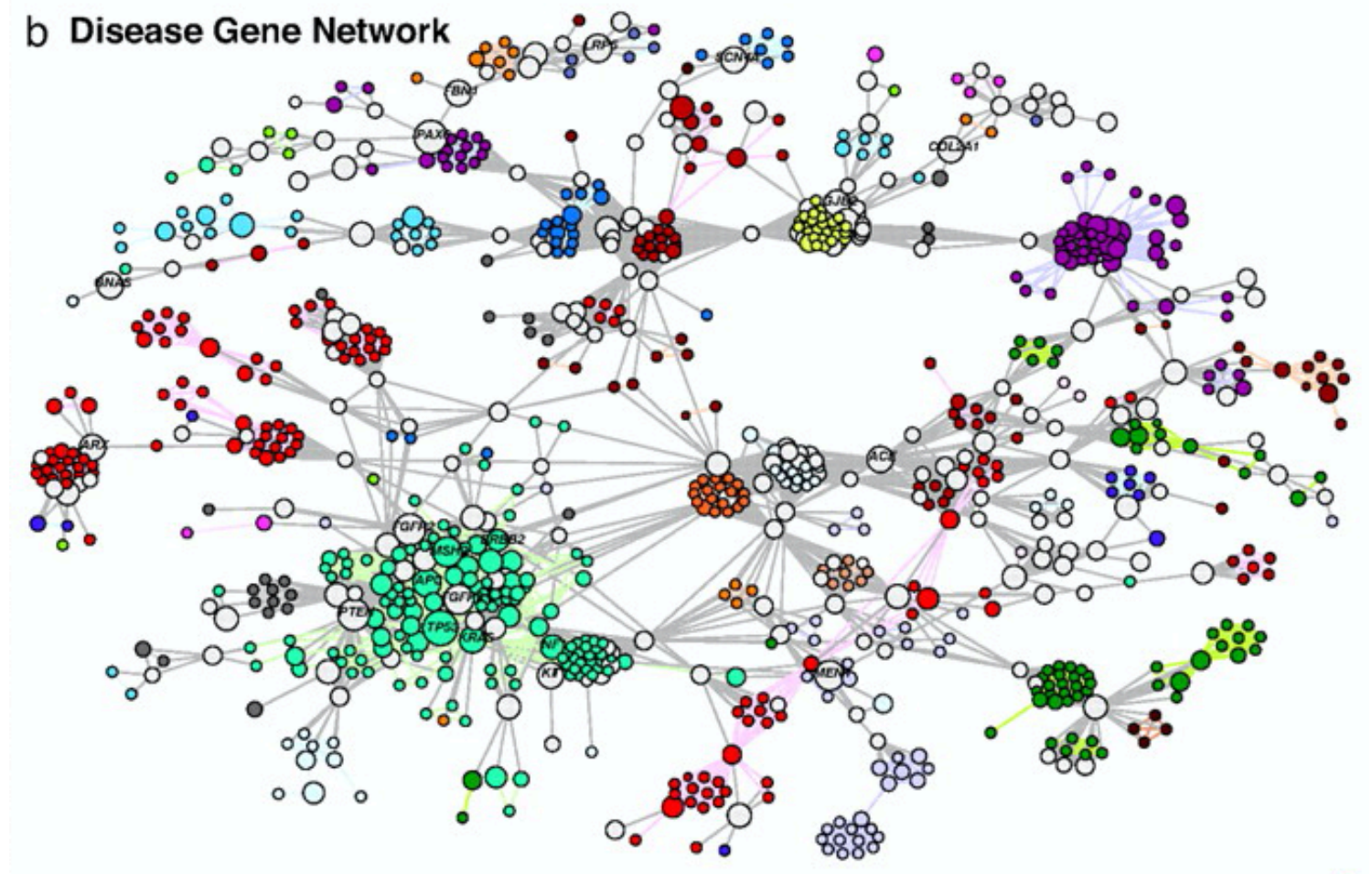
## ★ Human disease network

( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ THIS PROJECT (base):

Extract such networks directly from a set of GWAS



( from Goh et al PNAS '07: The human disease network )



# Project

## ★ Human disease network

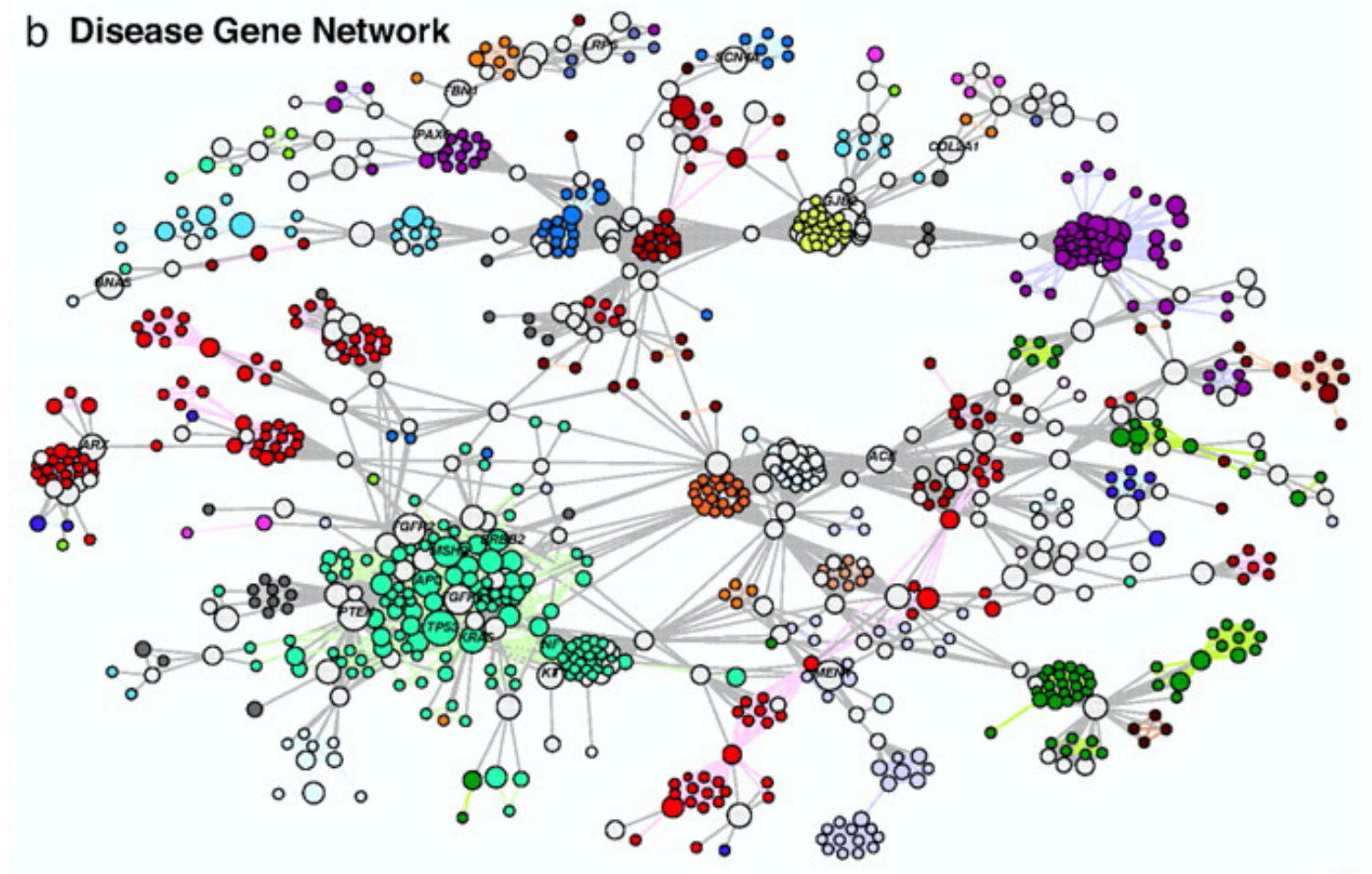
( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ THIS PROJECT (base):

Extract such networks directly from a set of

GWAS



( from Goh et al PNAS '07: The human disease network )

# GWAS

## ★ Genome wide association studies (GWAS)

In a nutshell: Linear regression of the genomic variations in a population (single nucleotide polymorphisms) onto traits, like for instance BMI or disease states.



# GWAS

## ★ Genome wide association studies (GWAS)

In a nutshell: Linear regression of the genomic variations in a population (single nucleotide polymorphisms) onto traits, like for instance BMI or disease states.

➡ By now summary statistics for a large quantity of traits are publicly available

# Project

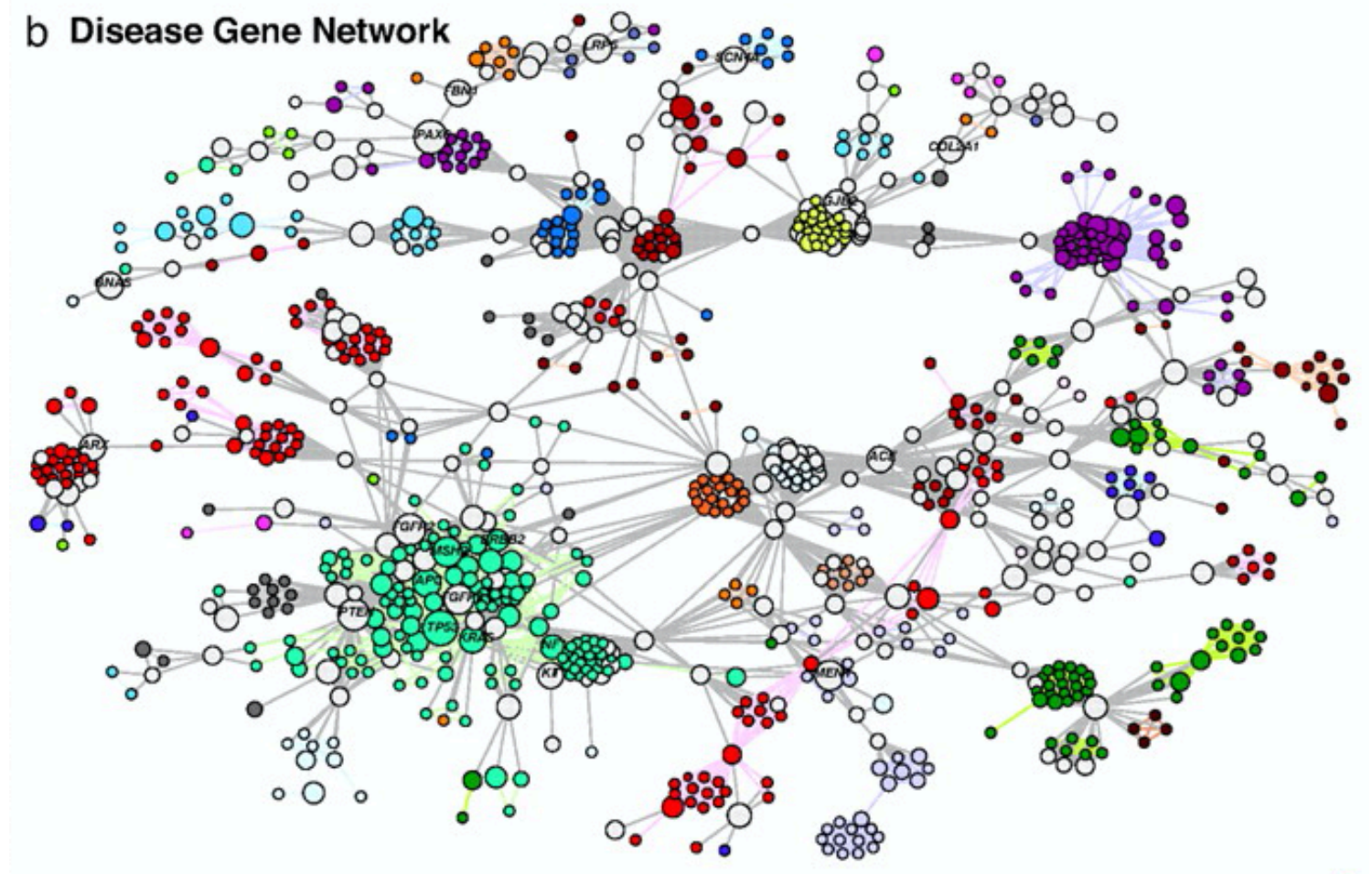
## ★ Human disease network

( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ THIS PROJECT (base):

Extract such networks directly from a set of GWAS



( from Goh et al PNAS '07: The human disease network )



# Project

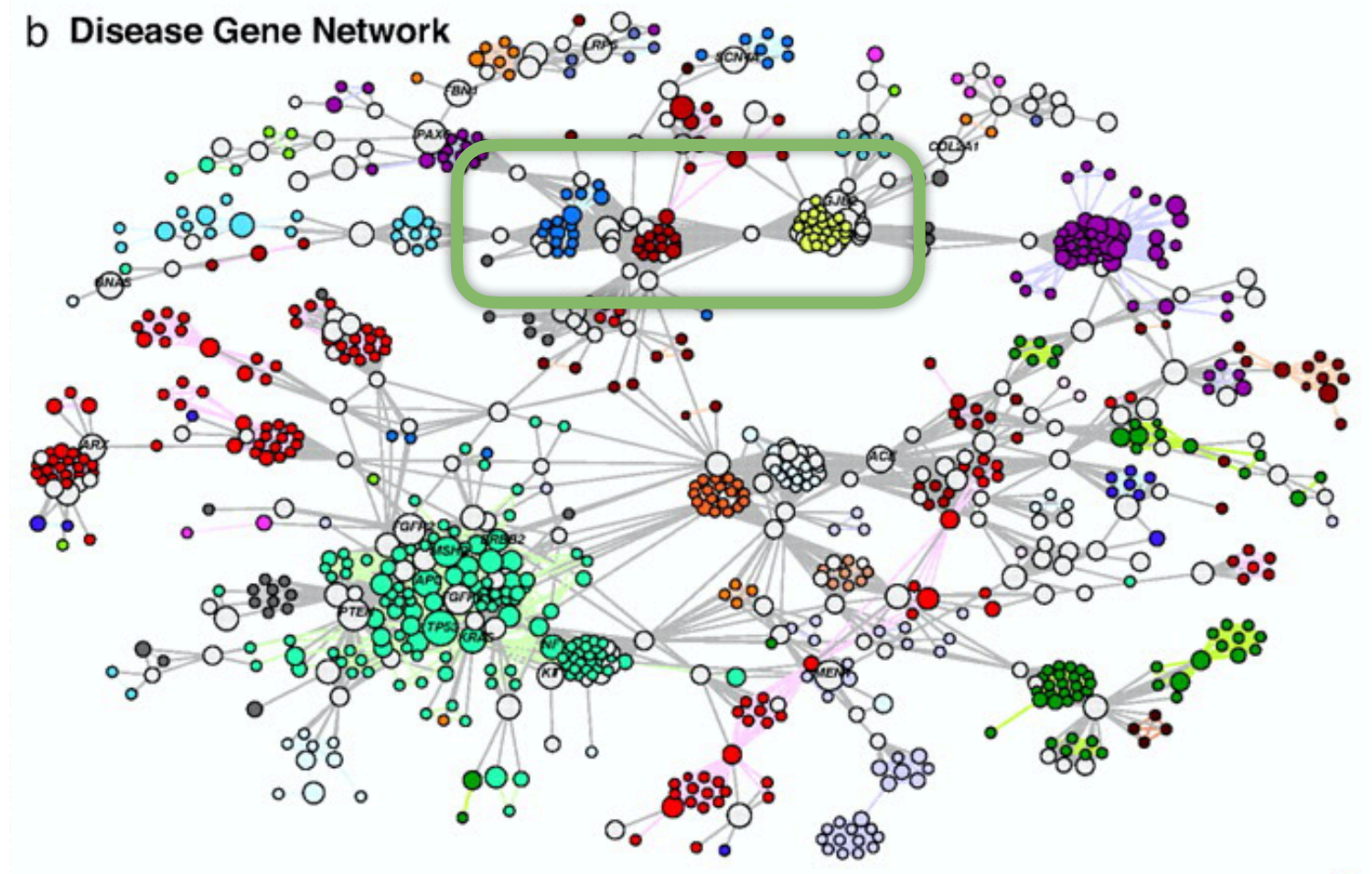
## ★ Human disease network

( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ THIS PROJECT (base):

Extract (small) networks directly from a set of GWAS



( from Goh et al PNAS '07: The human disease network )

# Project

## ★ Human disease network

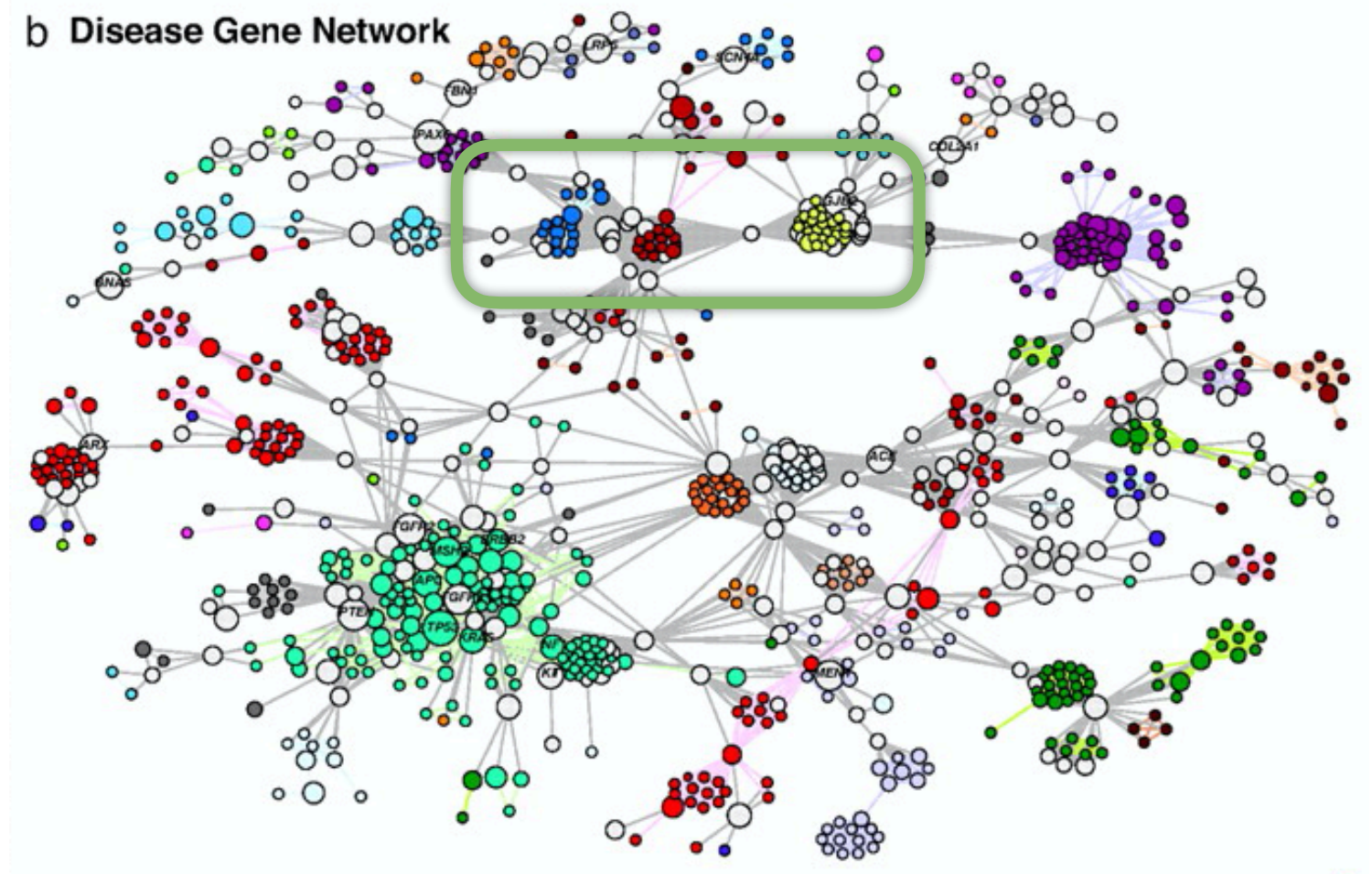
( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ THIS PROJECT (base):

Human-disease network:

Connect two traits (nodes)  
with edge weight  
proportional to the # of  
co-significant genes



( from Goh et al PNAS '07: The human disease network )



# Project

## ★ Human disease network

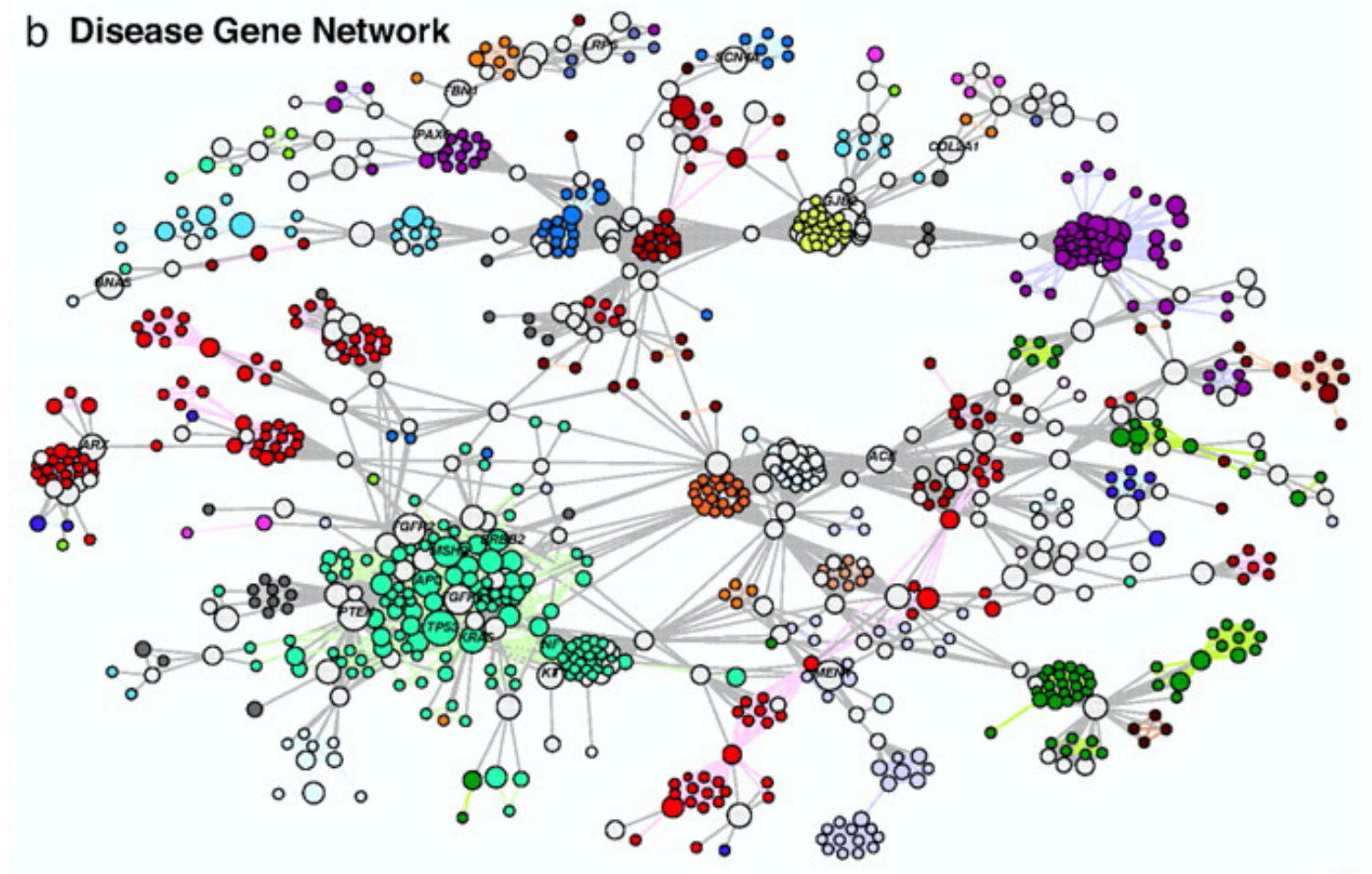
( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ THIS PROJECT (base):

Human-disease network:

Connect two traits (nodes)  
with edge weight  
proportional to the # of  
co-significant genes



( from Goh et al PNAS '07: The human disease network )

# Project

## ★ Human disease network

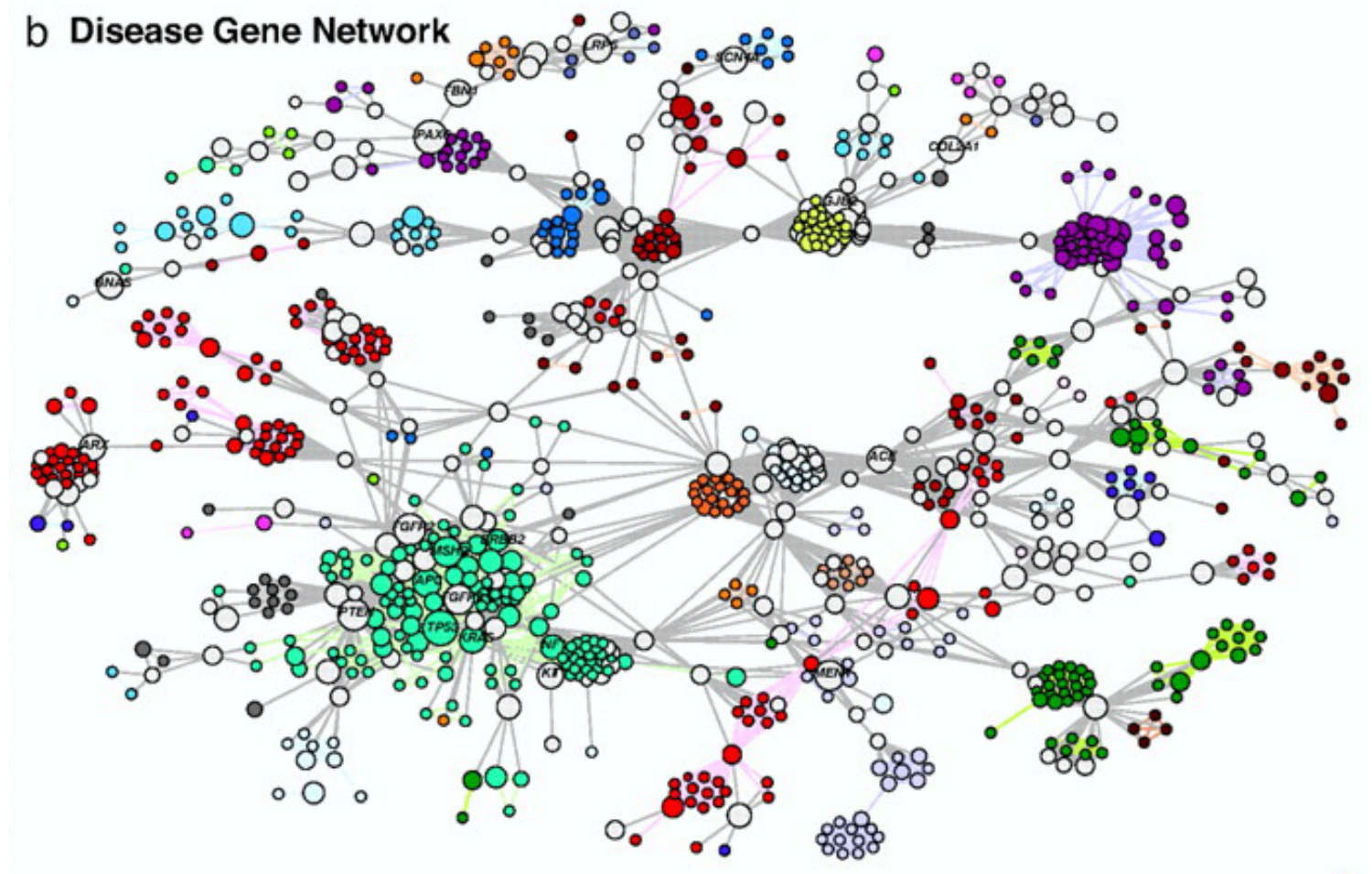
( Goh et al '07: The human disease network )

Via harvesting known Gene-Disease relations networks can be constructed and clustered:

## ★ THIS PROJECT (base):

Disease-gene network:

Connect two genes (nodes) with edge weight proportional to the # of traits the two genes are co-significant under



( from Goh et al PNAS '07: The human disease network )



# Outline

## ★ Step 1:

Collect suitable GWAS summary statistics from public resources

# Outline

## ★ Step 1:

Collect suitable GWAS summary statistics from public resources

## ★ Step 2a:

Use python package PascalX to construct (binary) gene co-significance matrices for all GWAS pairs (and protein coding genes)



# Outline

## ★ Step 1:

Collect suitable GWAS summary statistics from public resources

## ★ Step 2a:

Use python package PascalX to construct (binary) gene co-significance matrices for all GWAS pairs (and protein coding genes)

## ★ Step 2b:

Investigate pair-wise co-significances. Can you uncover new shared genes ?

# Outline

## ★ Step 1:

Collect suitable GWAS summary statistics from public resources

## ★ Step 2a:

Use python package PascalX to construct (binary) gene co-significance matrices for all GWAS pairs (and protein coding genes)

## ★ Step 2b:

Investigate pair-wise co-significances. Can you uncover new shared genes ?

## ★ Step 3a:

Integrate the pair-wise matrices to a Human-disease network.



# Outline

## ★ Step 1:

Collect suitable GWAS summary statistics from public resources

## ★ Step 2a:

Use python package PascalX to construct (binary) gene co-significance matrices for all GWAS pairs (and protein coding genes)

## ★ Step 2b:

Investigate pair-wise co-significances. Can you uncover new shared genes ?

## ★ Step 3a:

Integrate the pair-wise matrices to a Human-disease network.

## ★ Step 3b:

Open end analysis of the network ...

# Outline

## ★ Step 1:

Collect suitable GWAS summary statistics from public resources

## ★ Step 2a:

Use python package PascalX to construct (binary) gene co-significance matrices for all GWAS pairs (and protein coding genes)

BASE

## ★ Step 2b:

Investigate pair-wise co-significances. Can you uncover new shared genes ?

## ★ Step 3a:

Integrate the pair-wise matrices to a Human-disease network.

## ★ Step 3b:

Open end analysis of the network ...



# Technicalities

## ★ Data:

Publicly available GWAS summary statistics.

( GWAS catalog: <https://www.ebi.ac.uk/gwas/downloads/summary-statistics> )

## ★ Tools:

**Python** ecosystem:

PascalX, jupyter, numpy, scipy, scikit-learn, pandas, matplotlib, ...

( Prior knowledge of Python recommended )

...

# Outcome



You will acquire:

- General knowledge about GWAS studies
- Experience working with large scale GWAS summary statistics
- Data pre-processing skills in python and linux shell
- Big data skills in python
- General knowledge about the spectrum of human diseases and their interplay



Feel free to contact me by email if you  
have any questions !

[daniel.krefl@unil.ch](mailto:daniel.krefl@unil.ch)